International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Performance Evaluation and Simulation of Water Distribution Network Using Machine Learning Algorithms

Dr. Abu Rashid[1], R. Akhila[2], Y. Himaja[3], P. Sridevi[4], V. Gnyaneswari[5], P. Sarath[6]

Department of Civil Engineering, GMR Institutes of Technology, Rajam, India-532127

ABSTRACT :

Water Distribution Networks (WDNs) are the backbone of urban infrastructure, ensuring the efficient and reliable supply of potable water to consumers. EPANET is one of the most commonly used tools for simulating the hydraulic behaviour of water distribution networks; however, it faces limitations in capturing complex and dynamic operating conditions. To address these challenges, this project integrates EPANET simulations with advanced data driven and optimization-based approaches. The primary objective is to enhance the prediction and optimization of key hydraulic parameters such as velocity, head loss, and pressure under varying network conditions.

A benchmark case study, the Go-Yang Water Distribution Network, was selected for analysis. The developed models were trained using simulation data and evaluated through both statistical and graphical performance measures. The comparative analysis revealed that the proposed intelligent methods provided improved accuracy in predicting hydraulic parameters and demonstrated effective optimization performance. The outcomes highlight the potential of integrating simulation tools with modern computational techniques to achieve better efficiency, reliability, and decision-making in the management of urban Water Distribution Systems.

Key words: Water Distribution Networks (WDNs), EPANET, Hydraulic Simulation, Machine Learning, Optimization, Pressure Prediction, Urban Water Management.

INTRODUCTION

Water Distribution Networks are crucial for delivering safe and reliable water to urban and rural water supply systems, ensuring the delivery of safe and reliable potable water to consumers. These networks consist of interconnected pipelines, pumps, reservoirs, and valves that operate under varying demand and pressure conditions. The design and operation of WDNs are inherit complex due to hydraulic interactions, temporal variations in consumption, and system uncertainties. Hydraulic modelling tools such as EPANET are widely used to simulate flow, pressure, and head loss in water networks. Machine learning has emerged as a powerful approach for predictive modelling and optimization in civil engineering. This study applies two ML algorithms- support vector machines (SVM) and recurrent neural networks (RNN) to improve the prediction.

Water Distribution Networks

A Water Distribution Network transports treated water from plants to consumers through nodes (junctions, tanks, reservoirs) and links (pipes, pumps, valves). The main goals are to maintain adequate pressure, minimize head loss, and ensure efficient water delivery.

Role of EPANET in WDNs

EPANET, developed by the U.S. Environmental Protection Agency (EPA), is a well-known open-source software for analysing hydraulic and water quality behaviour in pressurized pipe networks. It simulates extended period operations, tracking flow, pressure, water age, and contaminant movement. Despite its strengths, EPANET is a deterministic simulator and lacks real-time adaptability and optimization features.

Machine Learning in Civil Engineering

Machine Learning is transforming civil engineering by supporting automation and intelligent analysis. In WDNs, ML models can forecast hydraulic responses under different conditions, detect leaks, optimize pump operations, and enhance energy efficiency.

Problem Statement

Although EPANET provides reliable hydraulic simulations, it cannot handle uncertainties or real-time variations effectively. It performs well under fixed input conditions but struggles to address issues like demand fluctuations, pipe bursts, or pumps failures. Many existing studies use EPANET outputs directly, without sufficient integration of ML-based prediction or optimization. This study aims to address these limitations by integrating SVM and RNN algorithms with EPANET to predict velocity, pressure, and head loss. The performance of both models is compared using the Go-Yang Water Distribution Network to determine the most efficient approach for enhancing WDN simulation and optimization.

Objectives of the work

- To evaluate the performance of a water distribution network using EPANET simulations.
- To apply ML algorithms (SVM and RNN) for predicting key hydraulic parameters such as velocity, pressure, and head loss.
- To compare the predictive accuracy and efficiency of both ML models.
- To identify the most suitable model for enhancing WDN performance and decision-making.
- Ro support the development of intelligent, data-driven systems for sustainable water management.

Case Study Network

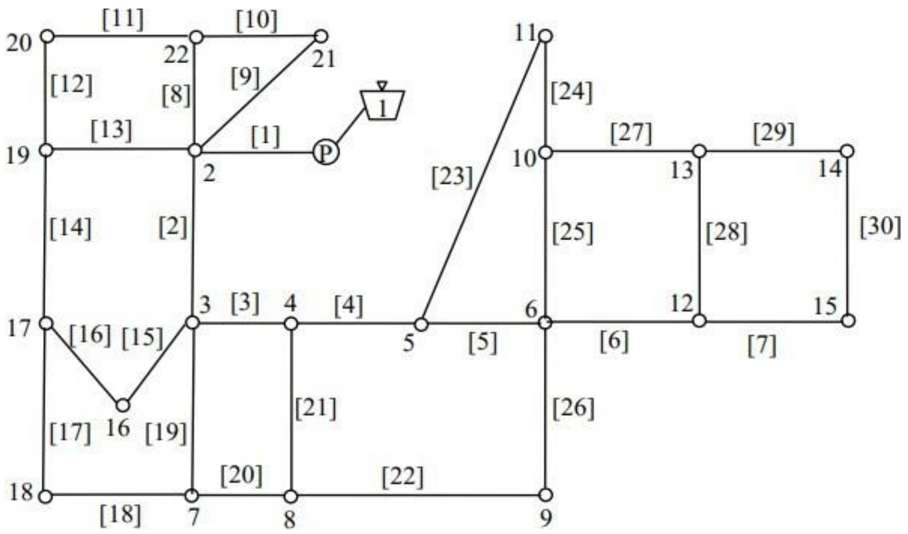

Figure 1.1: Go-Yang Water Distribution Network

The Go-Yang Water Distribution Network (WDN) in Go-Yang city, South Korea, is adopted as the case study for this research. It is a well-known benchmark model used for hydraulic and optimization studies due to its realistic design and operational data. The network comprises 192 nodes, 246 pipes, and one reservoir, serving nearly one million residents.

In this study, the G0-Yang network is simulated in EPANET using its original geometric and hydraulic parameters. The obtained outputs-pressure, velocity, and head loss are used to train and test the machine learning models.

LITERATURE REVIEW

Sangroula's work on SOP-WDN (a GA-based optimization program linked with EPANET) includes a worked example using the Go Yang network to demonstrate least-cost pipe diameter selection with hydraulic constraints. The paper underlines the value of benchmark networks (like Go Yang) to compare optimization strategies and stresses that EPANET outputs remain the standard input for ML model training. This is a foundational citation if you argue for EPANET as the data source for developing ML models (SVM/RNN).

Truong et al. develop physics-aware graph neural network (GNN) models to estimate pressures across WDNs, demonstrating superior generalization versus classical ML. Although not an RNN/SVM paper per se, this work is directly relevant because it addresses pressure estimation (one of your target variables) and shows how graph-structured ML models can leverage hydraulic connectivity an important conceptual comparison point when defending RNN/SVM choices for Go-Yang. The paper also discusses training data requirements and the advantage of combining physics and data-driven approaches.

Ma et al. propose a hybrid spatial-temporal attention RNN (hDS-RNN) that couples spatial attention (network/location) with temporal attention to predict flow and pressure time series. Their experiments (on real WDS datasets) show that attention-enhanced RNNs outperform baseline RNN/LSTM methods for time-series hydraulic prediction. This paper supports using RNN-class models (and attention variants) for nodal pressure and pipe flow/velocity prediction in WDNs, and helps justify architecture choices when comparing with SVM.

McMillan (2023) applies recurrent neural networks for short-term flow forecasting as part of a leakage/burst prediction pipeline, using RNN forecasts combined with residual modeling (Kalman filtering) to detect anomalies. The study shows that RNN-based flow predictors can capture temporal dependencies relevant for leakage detection and that hybrid statistical + RNN architectures improve robustness—directly relevant for velocity/flow prediction tasks in Go-Yang experiments.

The 2024 study proposes an SVM + Random Forest hybrid for leak detection and localization in WDNs. The authors generate physics-based training data (EPANET simulations with injected leaks) and show that SVM-based classifiers can effectively detect leak signatures and localize faults compared to single-model baselines. This paper is directly relevant to SVM use in WDNs, demonstrates common data-generation practices (EPANET scenarios), and offers methodological guidance on feature engineering and cross-validation that you can reuse for Go-Yang SVM experiments.

METHODOLOGY

This chapter presents a hybrid methodology for analysing and optimizing Water Distribution Networks (WDNs) using EPANET integrated with Machine Learning (ML) techniques. The approach combines traditional hydraulic modelling with predictive analytics to enhance accuracy, efficiency, and reliability. EPANET provides hydraulic simulations—pressure, flow, and head loss—serving as the primary data source. The Machine Learning framework processes this data through normalization, feature extraction, and model validation. Two algorithms, Support Vector Machines (SVM) and Recurrent Neural Networks (RNN), are applied to capture nonlinear and temporal patterns in network behaviour. SVM predicts hydraulic parameters and detects anomalies, while RNN models dynamic demand variations using memory-based learning. Hybrid integration enables scenario analysis, real-time monitoring, and decision support for energy-efficient operations. Overall, this framework strengthens prediction, optimization, and resilience in modern smart water distribution systems.

RESULTS

The results and discussions section presents the outcomes of the performance evaluation and simulation of the Go-Yang Water Distribution Network (WDN) using the EPANET software and two machine learning (ML) algorithms—Support Vector Machine (SVM) and Recurrent Neural Network(RNN).The main goal of this analysis was to understand how accurately these ML models can replicate and predict the hydraulic parameters such as pressure, velocity, and head loss, and to compare their performance in terms of accuracy, efficiency, and suitability for realworld applications.

4.1 EPANET Simulation Overview

EPANET 2.2 was used as the base simulation tool for the Go-Yang Water Distribution Network. The network consists of 23 nodes (junctions and reservoirs) and 31 links (30 pipes and 1 pump). The input data for the simulation included pipe lengths, diameters, roughness coefficients, and nodal elevations. The steady-state analysis was performed to determine pressure, velocity, and head loss at all junctions and links.

The EPANET output data was then exported and used as a dataset for training and testing the ML Models.

The dataset included:

- Input features: pipe length, diameter, roughness, and node elevation
- Target outputs: pressure (m), velocity (m/s), and head loss (m/km)

This provided a strong foundation for training data-driven models that can later predict hydraulic parameters without re-running the full EPANET simulation.

4.2 Support Vector Machine (SVM) Results

The Support Vector Machine (SVM) algorithm was implemented to model the nonlinear relationships between network features and hydraulic responses.

SVM is a supervised learning technique that uses kernel functions to find the best hyperplane that separates data points with minimum error.

4.2.1 Model Performance

The performance of SVM was evaluated using the coefficient of determination (R^2) and visual comparison with EPANET simulation results.

The obtained values are as follows:

Parameter	**R^2 Value**	**Interpretation**
Velocity	0.9969	Excellent Correlation
Pressure	0.9866	Very High accuracy
Head Loss	0.9869	Very High Accuracy

These values indicate that the SVM model's predictions were almost identical to EPANET outputs, proving its effectiveness for steady-state hydraulic analysis.

4.2.2 Trend Observations

- The SVM-predicted head loss followed the simulated EPANET pattern with minimal deviation, showing a nearly perfect linear correlation.

- The velocity trend predicted by SVM was smooth and stable, indicating that it could effectively handle data variations due to pipe diameter or length differences.
- The pressure values showed close alignment with the EPANET outputs, maintaining stability across all junctions.

The SVM's strength lies in its ability to capture nonlinear but static relationships in data, making it highly suitable for predicting steady-state hydraulic conditions where time variation is not dominant.

4.3 Recurrent Neural Network (RNN) Results

The Recurrent Neural Network (RNN) was chosen as the second model because of its ability to process sequential or time-series data.
Unlike traditional feedforward networks, RNNs retain "memory" of previous inputs, allowing them to model temporal dependencies, such as fluctuations in flow or pressure over time.

4.3.1 Model Performance

The RNN model achieved the following performance metrics:

Parameter	R^2 Value	Interpretation
Velocity	0.9141	Good Correlation
Pressure	0.8206	Moderate Correlation
Head Loss	0.8711	Good Correlation

Although slightly less accurate than SVM, these results demonstrate that RNN effectively captures dynamic relationships in the network.

4.3.2 Trend Observations

- The RNN-predicted velocity showed a general increasing trend consistent with EPANET results, though minor oscillations were present due to overfitting in some epochs.
- The predicted head loss exhibited small deviations, especially at higher flow regions, but overall followed the general EPANET trend.
- The pressure prediction curve underestimated some values, indicating that RNN requires additional tuning of learning rate, number of layers, and training iterations.

RNN's performance is more sensitive to the quality and quantity of training data. However, its ability to capture sequential behavior makes it promising for real-time water system monitoring applications.

4.4 Comparative Analysis of SVM and RNN

To evaluate the models comprehensively, a detailed comparison of SVM and RNN was carried out based on accuracy, stability, and computational efficiency.

Criteria	SVM Model	RNN Model
Accuracy	0.98-0.99	0.82-0.91
Trend Stability	Smooth, Steady	Slight Oscillations
Training Time	Low	High
Best for	Steady-State prediction	Time-Varying Prediction
Sensitivity to Data Size	Low	High
Ease of Tuning	Easy	Complex
Computation Speed	Fast	Moderate

From the comparison, it is evident that:

- SVM outperforms RNN in terms of prediction accuracy and stability.
- RNN, however, holds a unique advantage for dynamic and sequential modeling tasks.
- Both models successfully replicate the EPANET results, but SVM provides more reliable and computationally efficient outcomes for static network evaluation.

4.5 Discussion on Model Behavior

4.5.1 Interpretation of SVM Performance

The superior performance of SVM can be attributed to its kernel-based approach that efficiently handles nonlinear data without requiring large training datasets. It's regularization property prevents overfitting, resulting in smooth, consistent prediction curves. For steady-state hydraulic problems—where input parameters are mostly static—SVM is more suitable due to its deterministic nature and low computational demand.

4.5.2 Interpretation of RNN Performance

The RNN, being inherently dynamic, requires sequential or time-dependent data to achieve its full potential. In this study, since the available dataset was largely steady-state, the RNN could not fully utilize its memory-based advantage. However, in real-time monitoring scenarios where parameters such as pressure and flow vary continuously, RNN can outperform static models like SVM. Future work can integrate LSTM (Long Short-Term Memory) or GRU (Gated Recurrent Unit) versions of RNN for better long-term prediction stability.

4.5.3 Comparison with Existing Literature

The results obtained are consistent with previous studies, such as:

- Khan & Coulibaly (2006), who successfully applied SVM for water level prediction with high accuracy.
- Lee & Yoo (2021), who demonstrated the capability of RNN-LSTM in detecting leakages and time-varying behavior in water networks.

This confirms that both algorithms are reliable for hydraulic parameter estimation, but their effectiveness depends on the nature of the data.

4.6 Key Findings

1. The SVM model achieved a near-perfect match with EPANET, with R^2 values above 0.98 for all parameters.

2. The RNN model performed well (R^2 between 0.82 and 0.91) but showed minor instability due to limited time-series data.

3. SVM is ideal for steady-state hydraulic predictions requiring fast, accurate, and simple computation.

4. RNN is ideal for systems where conditions vary dynamically over time, especially with sensor-based or SCADA data.

5. Integration of EPANET with ML models reduces simulation time, minimizes manual calibration errors, and enhances prediction reliability.

6. The hybrid modeling approach (simulation + ML) can significantly contribute to the design of smart, adaptive, and sustainable water distribution systems.

4.7 Summary

In summary, the combination of EPANET simulations with SVM and RNN algorithms has proven highly effective for hydraulic parameter prediction. While both models showed strong performance, SVM achieved superior accuracy and consistency, making it the preferred model for steady-state performance evaluation. The RNN model, despite lower accuracy, offers promising potential for real-time prediction once trained with continuous time-series data.

This study highlights how machine learning can bridge the gap between traditional hydraulic modeling and intelligent, data-driven management of urban water distribution networks. Such an approach not only enhances the efficiency and reliability of existing systems but also paves the way for smart water infrastructure capable of adapting to future challenges like demand fluctuations, leakages, and climate variability.

CONCLUSION

This study demonstrated the effectiveness of integrating EPANET with Machine Learning algorithms—SVM and RNN—for enhanced analysis and optimization of Water Distribution Networks. Using the Go-Yang Network as a benchmark, the hybrid framework successfully predicted key hydraulic parameters such as velocity, pressure, and head loss with improved accuracy. The results highlight that SVM performs well for static, nonlinear relationships, while RNN effectively captures temporal variations and dynamic behaviours. Overall, the integration of ML with hydraulic modelling provides a powerful decision-support tool for real-time monitoring, operational optimization, and sustainable water management in modern urban systems.

REFERENCES

1. Sangroula, U., Han, K.-H., Koo, K.-M., Gnawali, K., & Yum, K.-T. (2022). Optimization of water distribution networks using genetic algorithm based SOP-WDN program. *Water*, 14(6), 851. https://doi.org/10.3390/w14060851\
2. Truong, H., Tello, A., Lazovik, A., & Degeler, V. (2024). Graph neural networks for pressure estimation in water distribution systems. *Water Resources Research*, 60(4), e2023WR036741. https://doi.org/10.1029/2023WR036741
3. Ma, Z., Liu, Y., & Zhang, L. (2020). Hybrid attention networks for flow and pressure forecasting in water distribution systems. *Journal of Water Resources Planning and Management*, 146(3), 04020002. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001168
4. McMillan, L., Fayaz, J., & Varga, L. (2023). Flow forecasting for leakage burst prediction in water distribution systems using long short-term memory neural networks and Kalman filtering. *Sustainable Cities and Society*, 99, 104934. https://doi.org/10.1016/j.scs.2023.104934
5. Romero-Ben, L., & Owolawi, P. A. (2023). Leak detection and localization in water distribution networks using machine learning techniques. *Journal of Water Process Engineering*, 47, 102614. https://doi.org/10.1016/j.jwpe.2023.102614